

Site :  Luminy  St-Charles  St-Jérôme  Cht-Gombert  Aix-Montperrin  Aubagne-SATIS

Sujet de :  1<sup>er</sup> semestre  2<sup>ème</sup> semestre  Session 2 Durée de l'épreuve : 2h

Examen de : M1 Nom du diplôme : Mathématiques Appliquées, Statistique

Code du module : SMSCU34C Libellé du module : Données catégorielles : Exploration, Modélisation

Calculatrices autorisées : Oui Documents autorisés : Oui

### Examen du lundi 12 décembre, 9h00-11h00

*Tous les documents qui ont été distribués dans le cadre du cours, en version papier ou via AMeTICE, ainsi que vos notes manuscrites et calculatrice sont autorisés. Tout accès à un ordinateur, à un téléphone et à internet est interdit.*

L'indice de position sociale (IPS<sup>1</sup>) [fourni](#) par le ministère de l'éducation nationale et de la jeunesse permet d'attribuer à chaque école élémentaire une valeur numérique dont l'objectif est de quantifier un ensemble d'attributs socio-économiques et culturels liés à la réussite scolaire des élèves de l'école. Dans la suite, nous nous restreindrons à l'ensemble des écoles élémentaires de France métropolitaine et à la variable qIPS qui est la version catégorisée de l'IPS suivant les quintiles. On trouvera en annexe A une carte des académies scolaires françaises ainsi que la définition de qIPS.

### Partie I

La table 1 présente pour chacune des académies de France métropolitaine le nombre d'écoles élémentaires suivant leur quintile d'appartenance à l'IPS. Ce tableau a été analysé de manière équivalente sous R et sous SAS. Les programmes et les sorties sont fournies en annexe B pour chaque logiciel. Vous répondrez aux questions en considérant une des deux annexes, à votre choix.

	[54.2,91.3[	[91.3,99.9[	[99.9,107.0[	[107.0,116.0[	[116.0,156.0]
AIX-MARSEILLE	301	166	169	251	274
AMIENS	407	273	224	168	111
BESANCON	119	172	157	183	87
BORDEAUX	220	387	408	344	311
CLERMONT-FERRAND	116	219	236	206	113
CORSE	50	47	41	19	16
CRETEIL	501	189	185	245	338
DIJON	177	242	240	207	119
GRENOBLE	228	252	400	544	527
LILLE	836	392	283	247	318
LIMOGES	65	96	120	93	26
LYON	277	211	264	350	444
MONTPELLIER	265	252	294	292	186
NANCY-METZ	311	313	258	224	156
NANTES	314	541	519	452	363
NICE	132	96	157	141	174
NORMANDIE	370	453	387	343	202
ORLEANS-TOURS	255	293	269	282	205
PARIS	46	39	39	57	276
POITIERS	148	295	223	193	116
REIMS	215	194	150	111	65
RENNES	230	450	487	440	343
STRASBOURG	119	113	144	213	199
TOULOUSE	193	314	388	429	355
VERSAILLES	335	178	168	264	912

Table 1: Nombre d'écoles élémentaires par académie et par modalité de la variable qIPS. Lecture : dans l'académie d'Aix-Marseille, 303 écoles élémentaires ont une valeur d'IPS comprise entre 54.2 et 91.3.

<sup>1</sup>Voir Rocher et al, 2016 pour la méthodologie de construction de l'indice.

1. Définir les deux nuages qui sont considérés dans l'analyse. Vous donnerez les coordonnées du centre de gravité du nuage relié aux académies ainsi que les points de ce nuage qui sont a) le plus proche du centre de gravité b) le plus éloigné du centre de gravité.
2. Dans les sorties du programme, quels éléments justifient d'analyser la structure de dépendance entre les variables associées au tableau présenté table 1 ?
3. Quelles sont les académies contribuant le plus au calcul de la distance du chi2 ? Que cela signifie-t-il dans l'étude considérée ?
4. En indiquant le(s) critère(s) choisi(s), combien d'axe(s) factoriel(s) proposeriez vous de retenir ?
5. Après avoir donné la valeur de la coordonnée de l'académie d'AIX-MARSEILLE sur l'axe 1, retrouver cette valeur en utilisant les relations quasi-barycentriques.
6. Donner la valeur du poids de la classe d'IQS [54.2, 91.3[ dans l'analyse.
7. Expliquer le calcul de la contribution relative au premier axe de la classe d'IQS [54.2, 91.3[.
8. Quelles sont les deux académies contribuant le plus à l'axe 1 ?
9. Donner une interprétation argumentée de chacun des deux premiers axes factoriels et de la représentation simultanée dans le plan 1-2 présentée dans l'annexe.

## Partie II

On veut étudier le lien entre le fait d'avoir un IQS supérieur ou égal à 116 et l'académie. La Table 2 donne le nombre d'écoles par académie ayant un IQS supérieur ou égal à 116, ou < 116. On note  $y_i$  la variable valant 1 si l'IQS de l'académie  $i$  est un IQS supérieur ou égal à 116, et 0 sinon.

Académie	Aix-Marseille	Amiens	Besançon	Bordeaux	Clermont	Corse	Créteil	Dijon	Grenoble	Lille	Limoges	Lyon	Montpellier
IQS < 116	887	1072	631	1359	777	157	1120	866	1424	1758	374	1102	1103
IQS ≥ 116	274	111	87	311	113	16	338	119	527	318	26	444	186

  

Académie	Nancy-Metz	Nantes	Nice	Normandie	Orléans-Tours	Paris	Poitiers	Reims	Rennes	Strasbourg	Toulouse	Versailles
IQS < 116	1106	1826	526	1553	1099	181	859	670	1607	589	1324	945
IQS ≥ 116	156	363	174	202	205	276	116	65	343	199	355	912

Table 2: Effectifs par académie des écoles ayant un IQS supérieur ou égal à 116.

10. Définir et calculer les estimations des quantités suivantes :
  - La cote d'avoir un IQS supérieur à 116 dans l'académie d'Aix-Marseille.
  - La cote d'avoir un IQS supérieur à 116 dans l'académie de Versailles.
  - Le rapport de cotes (odds-ratio) pour une école de l'académie de Versailles par rapport à une école de l'académie d'Aix-Marseille?
  - Interpréter les résultats précédents.
11. On choisit l'académie d'Aix-Marseille comme académie de référence. Définir le modèle de régression logistique permettant de modéliser la probabilité d'avoir un IQS supérieur à 116 en fonction de l'académie. Combien ce modèle a-t-il de paramètres?
12. L'estimation du maximum de vraisemblance du coefficient du modèle associé à l'académie de Versailles vaut 1.13917. Retrouver ce résultat à partir des questions précédentes en justifiant votre réponse.
13. Sachant que l'estimation de la variance de l'estimateur de ce coefficient vaut  $(0.08326)^2$ , tester l'hypothèse de nullité de ce coefficient par le test de Wald. On appellera les hypothèses à tester, la statistique utilisée et sa loi sous  $(H_0)$ . On rappelle que  $\mathbb{P}(|\mathcal{N}(0, 1)| > 1.96) = \mathbb{P}(\chi_1^2 > 3.84) = 0.05$ . Commenter le résultat obtenu.
14. Expliquez comment construire un intervalle de confiance de coefficient de sécurité 95% pour la probabilité qu'une école de l'académie de Versailles ait un IQS supérieur à 116.

# Annexe A

## 1 - Carte des académies de France métropolitaine



## 2 - Définition de la variable qIQS

qIQS est une variable catégorielle définie à partir de IQS de la manière suivante : une école prend la modalité

$$[54.2, 91.3[ \quad \text{si } 54.2 \leq IQS < 91.3;$$

$$[91.3, 99.9[ \quad \text{si } 91.3 \leq IQS < 99.9;$$

$$[99.9, 107.0[ \quad \text{si } 99.9 \leq IQS < 107.0;$$

$$[107.0, 116.0[ \quad \text{si } 107.0 \leq IQS < 116.0;$$

$$[116.0, 156.0] \quad \text{si } 116.0 \leq IQS \leq 156.0.$$